Oqtans: A Galaxy-integrated Workflow for Quantitative Transcriptome Analysis from NGS Data

Presented by Sebastian J. Schultheiss and Géraldine Jean

The current revolution in sequencing technologies allows us to obtain a much more detailed picture of transcriptomes. Studying them under different conditions or in mutants will lead to a considerably improved understanding of the underlying mechanisms of gene expression and processing. An important prerequisite is to be able to accurately determine the full complement of RNA transcripts and to infer their abundance in the cell. However, the analysis is made considerably more difficult by various limitations and biases in next-generation sequencing (NGS) technologies.

We present a Machine Learning powered platform for quantitatively analyzing RNA-seq experiments. It is integrated in the easy-to-use Galaxy framework [1] and builds on recent methods developed on the Max Planck Campus in Tübingen (Germany) for NGS sequence analysis:

PALMapper [2] is a short read mapper which efficiently computes both unspliced and spliced alignments at high accuracy by taking advantage of base quality information and computational splice site predictions.

mTIM [3] is a machine learning-based transcript reconstruction method, which exploits features derived from RNA-seq read alignments and from computational splice sites predictions to infer the exon-intron structure of the corresponding transcripts.

rQuant [4] is a method based on quadratic programming that simultaneously estimates biases inherent in library preparation, sequencing, and read mapping and accurately determines the abundances of given transcripts.

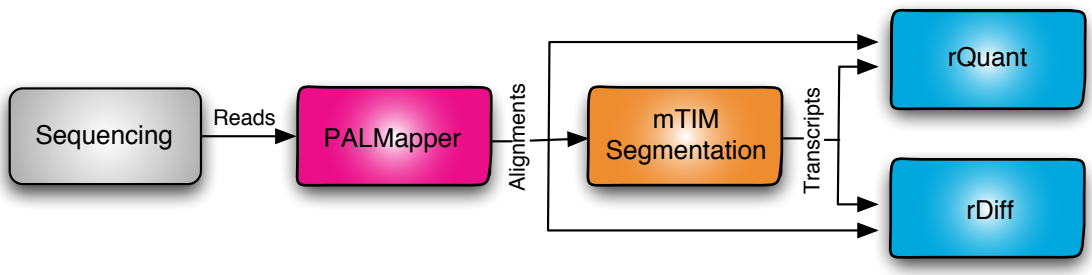
rDiff [5] is a set of statistical test techniques that determine significant differences between two RNA-seq experiments to find differentially expressed regions with or without knowledge of transcripts.

All these tools show very accurate results and perform better or on a par with the state-of-the-art for short read alignments, transcript identification and quantification as well as differential expression analysis.
Their combination into a powerful workflow integrated in the Galaxy framework makes it possible to easily and effectively conduct a complete quantitative RNA-Seq analysis. It can be accessed on our server, locally or in the cloud:

galaxy.fml.mpg.de/oqtans

Our programs are free, open source and standalone, and are available at:

oqtans.org



[1] D. Blankenberg *et al.*, Galaxy: a web-based genome analysis tool for experimentalists, *Curr Protoc Mol Biol*, 2010.

[2] G. Jean *et al.*, RNA-seq read alignments with PALMapper, *Curr Protoc Bioinform*, 2010.

[3] G. Zeller *et al.*, mTiM: margin-based transcript mapping from RNA-seq, *RECOMB-seq* 2011.

[4] R. Bohnert *et al.*, Transcript quantification with RNA-seq data, *BMC Bioinformatics*, 2009.

[5] O. Stegle *et al.*, Statistical tests for detecting differential RNA-transcript expression from read counts. *Nature Precedings*, 2010.

oqtans

galaxy.tuebingen.mpg.de
oqtans@tuebingen.mpg.de

Galaxy-Integrated Toolsuite
Machine-Learning-Powered

Quantitative RNA-Seq Analysis:
- Short Read Alignment
- Transcript Identification
- Transcript Quantification
- Differential Expression

oqtans developers

S. J. Schultheiss, G. Jean, J. Behr, R. Bohnert, P. Drewe, N. Goernitz, A. Kahles, P. Mudrakarta, V.T. Sreedharan, G. Zeller, G. Rätsch